Seguridad en IA: el enfoque de KPMG para una IA segura y escalable

KPMG. Make the Difference.

Marzo 2026

Asegurar la IA a gran escala

A medida que las organizaciones aceleran la adopción de la IA en procesos empresariales críticos, asegurar sistemas de IA a gran escala se ha convertido en un imperativo estratégico. A diferencia del software tradicional, la IA introduce **superficies de ataque únicas,** que van desde pipelines de datos y trAI de modelos hasta inferencia en tiempo real e interacción humano-IA. Estos sistemas dependen de vastos conjuntos de datos, arquitecturas de modelos complejas, dependencias externas y comportamientos dinámicos que evolucionan con el tiempo. Como resultado, las empresas deben proteger no solo su **infraestructura,** sino también la integridad de los **modelos,** la confidencialidad de los **datos** y la seguridad de **las decisiones impulsadas por IA.** Lograr una IA Segura a gran escala requiere un enfoque holístico y de todo el ciclo de vida que integre **seguridad, gobernanza, monitorización y protección de uso responsable** en cada capa del ecosistema de IA.

8 desafíos a los que se enfrentan nuestros clientes y nuestros expertos destacan

Descubrimiento e inventario de activos y sistemas de IA

Las organizaciones tienen dificultades para identificar técnicamente modelos, agentes, APIs y pipelines de IA desplegados en entornos cloud, on-premise y SaaS. Sin una visibilidad precisa, no se pueden aplicar controles de seguridad consistentes.

Entrenamiento y fine-tuning de modelos

Un reto crítico es **asegurar los datos usados para entrenar y ajustar modelos,** incluyendo controles sobre origen, integridad, calidad y permisos. La exposición a data **poisoning,** uso de datasets no autorizados o datos sensibles sin anonimizar es elevada, especialmente en entornos de IA generativa y aprendizaje continuo.

Protección de modelos para ataques específicos de IA

Los modelos de IA presentan **vectores de ataque distintos al software clásico,** como model extraction, inversion attacks, prompt injection, jailbreaks o manipulación de agentes. Muchas organizaciones no disponen de **controles técnicos específicos** para proteger modelos en reposo, en tránsito y en tiempo de inferencia.

Control de acceso y gestión de privilegios en MLOps

En plataformas de datos y MLOps suele existir **excesiva permisividad técnica:** desarrolladores con acceso directo a modelos productivos, falta de separación entre entornos (dev, test, prod) o credenciales compartidas para APIs de inferencia. El reto es implantar **IAM y RBAC adaptados a IA,** no heredados sin más del mundo IT tradicional.

Monitorización en tiempo de ejecución del comportamiento del modelo

Una vez desplegados, los modelos pueden **degradarse o comportarse de forma no esperada** debido a data drift, concept drift o cambios en el contexto de uso. Técnicamente, muchas organizaciones carecen de capacidades de **monitorización continua del comportamiento del modelo,** más allá de métricas básicas de rendimiento.

Trazabilidad técnica y versiones

Mantener **trazabilidad técnica completa** (qué versión de modelo, con qué datos, qué parámetros y en qué entorno) sigue siendo complejo. La falta de versionado riguroso dificulta auditorías técnicas, análisis forense tras incidentes y rollback seguro en caso de fallo del modelo.

Capacidades específicas de respuesta a incidentes de IA

La mayoría de las organizaciones no suelen tener **playbooks técnicos de respuesta a incidentes específicos de IA:** ¿qué hacer ante un modelo comprometido?, ¿cómo aislar un agente autónomo?, ¿cómo revocar un modelo sin impactar procesos críticos? La ausencia de estos mecanismos complica la contención y recuperación ante incidentes.

Modelos de terceros y riesgos de dependencia

La fuerte dependencia de modelos de fundación, bibliotecas de código abierto y servicios externos introduce riesgos técnicos opacos. Las dependencias de la IA a menudo no están cubiertas por los procesos tradicionales de análisis de composición de software.

Nuestra experiencia y enfoques de trabajo

Nuestro **portfolio de capacidades entorno a la seguridad de la IA,** se define como un roadmap de soporte para nuestros clientes encaminado a evaluar, establecer y aplicar marcos de gobierno, controles, modelos operativos y tecnológicos para lograr una IA segura y confiable.

01

Framework de control y operación del modelo

Desarrollo de un marco de controles de seguridad IA y operativizarlo para proteger los sistemas de IA durante todo suciclo de vida y anticipar riesgos desde el diseño.

Nuestro marco consolida más de 120 controles específicos respaldados por guías técnicas, e incorpora mecanismos estructurados para definir planes de acción y asignar responsabilidades.

02

Seguridad en la plataforma e infraestructura

Protección del entorno de IA aplicando controles alineados a riesgos para garantizar integridad, disponibilidad y escalabilidad de la plataforma e infraestructura.

Planteamos un enfoque integral para securizar la arquitectura e infraestructura de IA, asegurando controles adaptados en cada capa —desde la Landing Zone como base segura hasta la exposición de modelos.

03

Seguridad en el uso del modelo

Protección y supervisión del comportamiento del modelo mediante guardrAIls, trazabilidad y detección de anomalías para prevenir abusos y garantizar control en tiempo real.

Ofrecemos un servicio especializado para la protección, control y supervisión del comportamiento de modelos de IA, con un enfoque que combina medidas de seguridad en tiempo real con capacidades avanzadas de trazabilidad.

04

Testing y pruebas continuas de modelos

Evaluación de la resiliencia de los modelos frente a pruebas manuales y automatizadas, simulando ataques y condiciones críticas para garantizar su seguridad y robustez.

Nuestro enfoque se basa en la ejecución de pruebas automatizadas y manuales, adaptadas a los casos de uso, incluyendo la simulación de amenazas y vulnerabilidades reales, la evaluación de comportamiento o pruebas éticas y de robustez operativa.

05

Formación y concienciación

Sensibilización y capacitación sobre riesgos, oportunidades y buenas prácticas en el uso de la IA, abordando conceptos clave, casos de uso, medidas preventivas y guías de uso responsable.

Desarrollamos contenidos adaptados a distintos perfiles mediante formatos dinámicos —como campañas, simulaciones, guías y material educativo— y aplicamos estrategias de comunicación que maximizan la claridad, el alcance y la utilidad de la formación en IA.

Desde KPMG, adaptamos nuestros servicios a cada cliente y los prestamos de manera conjunta o por separado, con el objetivo de dar respuesta a la preocupaciones y casos de uso específicos, **en función del nivel de madurez** de cada compañía.

De este modo, **acompañamos a nuestros clientes en todo el ciclo de adopción segura de la IA** —desde las fases iniciales de evaluación y diagnóstico, hasta la gestión continua del riesgo y la mejora evolutiva— asegurando que cada paso esté **alineado con su contexto tecnológico, regulatorio y estratégico.**

De la evaluación a la acción: cómo ayudamos a mitigar los riesgos de la IA

Nivel 1

AI Security Foundation

La compañía está empezando a usar IA sin un enfoque estructurado de seguridad, con controles reactivos y dependientes de personas clave.

Establecemos los fundamentos: diagnóstico inicial, definición del marco básico de controles y diseño de un modelo organizativo mínimo viable.

Nivel 2

AI Security Advanced

La compañía ha puesto orden y control sobre los proyectos de IA más críticos, pero aún tiene una adopción desigual de controles y procedimientos.

Estandarizamos, formalizamos y extendemos los controles, integrarlos en los procesos corporativos y mejorar la supervisión y la trazabilidad.

Nivel 3

AI Security by Design

La compañía integra la seguridad de IA en su modelo operativo y de negocio, con controles preventivos, métricas y mejora continua.

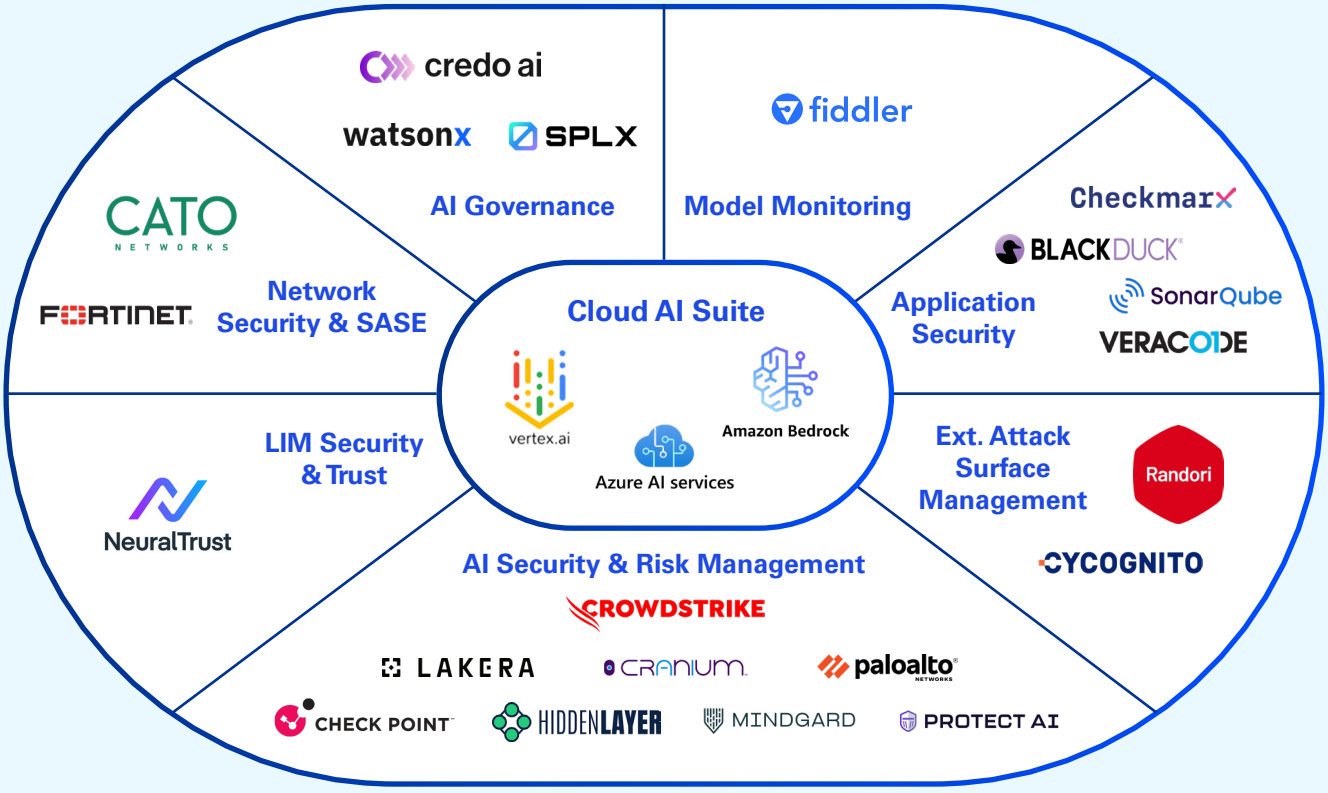
Consolidamos, fortalecemos y optimizamos la gobernanza, automatización, resiliencia y acompañamos en la evolución del modelo para escalar la IA de forma segura y sostenible.

Tecnología, experiencia y alcance global

La tecnología como palanca de securización

La tecnología es un medio indispensable para garantizar la seguridad y la confianza en la inteligencia artificial. A través de herramientas especializadas, desde KPMG conseguimos proteger modelos y datos, prevenir ataques y vulnerabilidades en entornos de IA y asegurar la integridad de todo el ciclo de vida de los sistemas.

El uso de **herramientas especializadas** de seguridad IA asegura una **protección integral** en todo el ciclo de vida de la inteligencia artificial. Estas soluciones permiten **blindar modelos y datos frente a ataques,** detectar y mitigar vulnerabilidades en tiempo real, proteger la cadena de suministro y garantizar la transparencia y confianza en cada fase, **desde el desarrollo hasta la operación** en entornos híbridos y multicloud.



Presencia internacional: nuestros clientes, referencias y metodologías nos avalan

Desde nuestra **posición de liderazgo global en AI Security,** estamos acompañando a organizaciones internacionales en la definición de metodologías, procesos y soluciones para una adopción segura de la inteligencia artificial.

Sector bancario

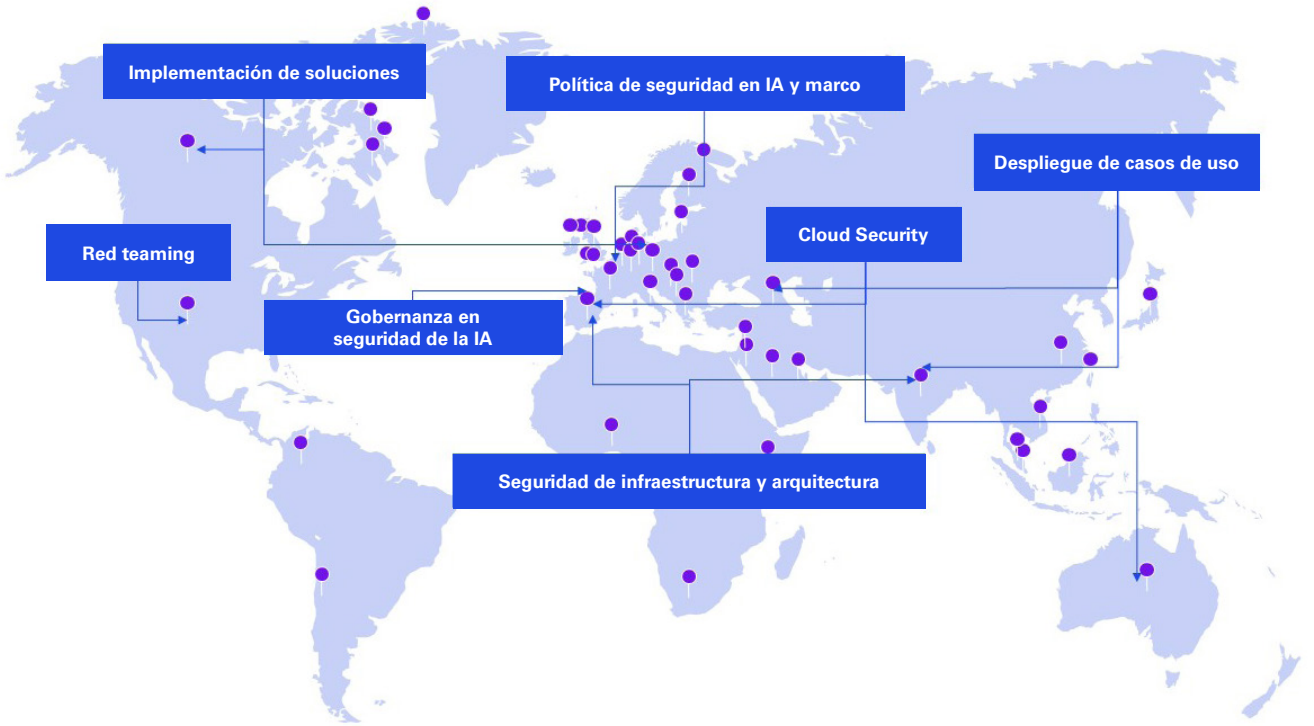
Creamos un marco de control y seguridad para IA, diagnosticando su situación, definiendo requisitos y diseñando un plan de implantación y mejora continua.

Sector retAIlr

Red Team sobre los chatbots de la compañía para identificar vulnerabilidades, y detectar brechas de configuración que permitían accesos indebidos a info. sensible.

Sector energético

Prestamos apoyo experto en seguridad e IA para identificar gaps, evaluar nuevas tecnologías y definir políticas y requisitos que fortalezcan su protección y gobierno.



Un equipo de expertos distribuido a lo largo de nuestra red global de firmas

Nuestro grupo de trabajo en seguridad de **IA se expande a +20 países con equipos multidisciplinares,** compartiendo activos para garantizar encubrimiento global, innovación continua, y un enfoque en seguridad de 360º para nuestros clientes.

Sea cual sea tu punto de partida, te ayudamos a proteger tus entornos de IA

Contacts

Global Leaders

Laurent Gobbi
Francia
Global Cyber & Tech Risk Leader
lgobbi@kpmg.fr

Jim Wilhelm
EE.UU
Global Cyber Investment Leader
jameswilhelm@kpmg.com

Javier Aznar
España
Global AI Security Leader
jaznar@kpmg.es

AI Security SMEs

Kristy Hornland
EE.UU
Director Advisory Cybersecurity & Tech Risk
khornland@kpmg.com

Katie Boswell
EE.UU
Managing Director Cybersecurity & Tech Risk
katieboswell@kpmg.com

Evan Rowell
EE.UU
Advisoring Managing Director
Consulting Managed Services
erowell@kpmg.com

Sydney Schemenauer
EE.UU
Manager Advisory Cybersecurity & Tech Risk
sschemenauer@kpmg.com

Mark Wilson
UK
Cyber Strategy - Senior Manager
Mark.WilsonRamsay@kpmg.co.uk

Markus Hupfauer
Alemania
FS Technology & IT-Compliance
mhupfauer@kpmg.com

Shirish Jangid
Oriente Medio
Director Digital Trust
sjangid1@kpmg.com

Gerry Chng
Singapur
Partner Cybersecurity & Tech Risk
gerrychng@kpmg.com.sg

Marta Pérez
España
Senior Manager AI Security PMO Coordinator
martaperezm@kpmg.es

Alexander Zagnetko
Eslovakia
Manager AI Security PMO Coordinator
alexanderzagnetko@kpmg.sk

KPMG

MAKE AI REAL

KPMG materializa el potencial de la IA en casos de uso reales que aceleran la generación de valor e impulsan la eficiencia.